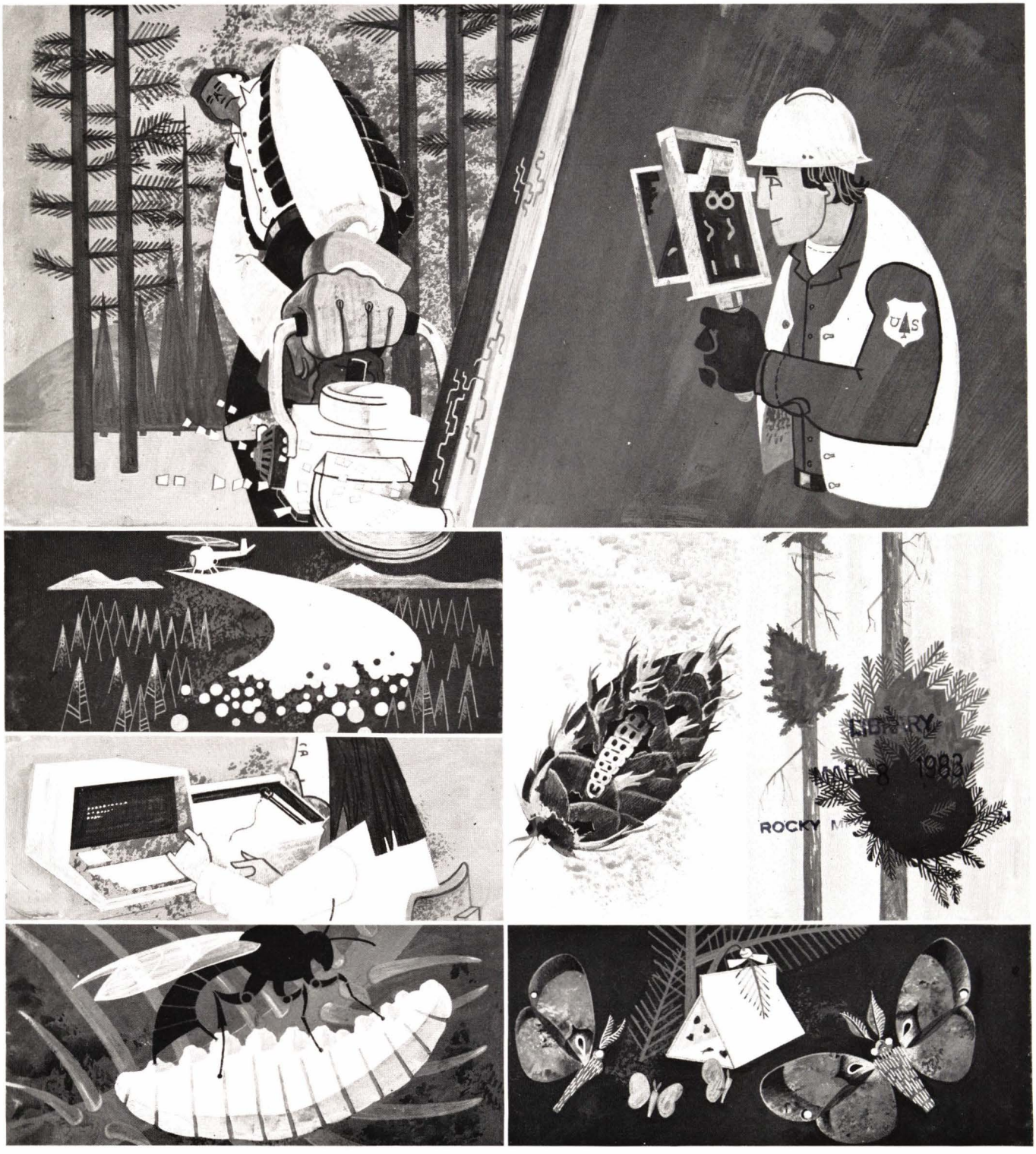
SB761
U8
#83-3
C.1

ed States
irtment of
culture
Forest
Service
Forest Pest Management
Methods Application Group
Fort Collins,
Colorado 80526

SAMPLING DESIGNS AND ALLOCATIONS YIELDING MINIMUM COST ESTIMATORS FOR MOUNTAIN PINE BEETLE LOSS ASSESSMENT SURVEYS



LIBRARY
MAR 8 1983
ROCKY M

Report No. 83-3

3400
November 1982

SAMPLING DESIGNS AND ALLOCATIONS YIELDING MINIMUM COST ESTIMATORS FOR MOUNTAIN PINE BEETLE LOSS ASSESSMENT SURVEYS

Nancy X. Sharpnack[1/] and John Wong[2/]
U.S.D.A. Forest Service

ABSTRACT

Tables of optimum sample size for each of three stages are presented for estimating mountain pine beetle loss in ponderosa and lodgepole pine forests in the western United States. These are listed for varying levels of precision and are based on data collected during surveys conducted between 1977 and 1980.

INTRODUCTION

Information from previous mountain pine beetle loss assessment surveys has provided the opportunity to improve future sampling designs and allocations. The material for this study was obtained from several surveys of ponderosa and lodgepole pine mortality that were conducted by the Forest Service in the western Regions between 1977 and 1980. These surveys provided a sufficient data base from which variance components were estimated, costs were assessed and various strategies could be evaluated.

The procedures used in the previous surveys and their results have been documented in various reports (Hostetler and Young 1979, Bennett and Bousfield 1978, Bennett et al. 1980, and Lister and Young 1981). The steps involved:

1. Aerial sketchmapping.
2. Stratification based on intensity of mortality per acre as determined from the sketchmapping.
3. Random sample of aerial photos within each stratum (Stage 1).
4. Subsample of aerial photos chosen with probabilities proportional to photo interpreted dead tree counts (Stage 2).
5. Sample of ground plots selected with probabilities proportional to dead tree counts (Stage 3).

---

1 Mathematical Statistician, Pacific Southwest Forest and Range Experiment Station, Berkeley, California.

2 Formerly Mathematical Statistician, Forest Pest Management, Methods Application Group, Davis, California. Presently Mathematical Statistician, Information Systems Management Staff, Pacific Southwest Region, San Francisco, California.

The objectives of this study were to evaluate whether or not stratification is effective, and to find those allocations of sampling units between the three stages which costs the least for a fixed percent standard error. The 20% standard error of the estimate is a requirement specified in the Forest Insect and Disease Information System Implementation Plan (FIDIS) (Ciesla and Yasinski 1980). We also considered the optimum allocation for 10%, 15% and 25% standard errors whenever these were attainable. We were constrained to considering 40 and 90 acre photo plots and 2.5 acre ground plots since all of the usable data available from the previous surveys fell into these categories. We considered the effects of taking one, two or three ground plots per photo plot in the third sampling stage where traditionally two plots have been taken.

METHODS

We used data from four previous mountain pine beetle surveys (Table 1). Cost factors applicable to the analysis were obtained from data furnished by Dayle Bennett (personal communication) in Region 3 and confirmed by Richard Myhre of the Rocky Mountain Forest and Range Experiment Station, Ft. Collins, Colorado. These costs are in units of the number of person hours to do photo interpretation and obtain ground measurements (Table 2). The cost estimates include plot set-up time and travel time. As expected, the time required to make a ground measurement is significantly higher than to interpret a photo, so that it is evident that a major cost element in sampling is associated with the time spent on the ground.

For each of the previous surveys, the variance components associated with each of the three stages were computed. From these it was possible to estimate the variance which would have resulted for any allocation of sampling units of interest. The estimate of total variance was computed using the following formula:

$$v(y) = (N-n/N)(1/n)v(y_1) + (1/m)v(y_2) + (1/ml)v(y_3)$$

in which

y = the estimate of total mortality

$v(y)$ = the estimate of the variance of the estimate of total mortality

$v(y_i)$ = the estimate of stage i variance

N = the total number of possible photo plots in the stratum

n = the number of photo plots to be sampled at Stage 1

m = the number of photo plots to be subsampled at Stage 2

l = the number of ground plots per photo plot at Stage 3

TABLE 1. Summary of Mountain Pine Beetle data sets used in generating estimates of variances for this study.

SURVEY	TREE SPECIES	STRATUM		:number of photos (Stage 1)	:photo plot size (acres)	:number photos subsampled (Stage 2)
1978 Beaverhead/ Gallatin	LP	L	< 4.9 trees/acre	94	40	24
		M	5.0-9.9	24		16
		H	≥ 10.0	26		16
1979 Montana	LP	L	<10.0 trees/acre	174	40	18
		M	10.0-41.0	87		19
		H	>41.0	81		20
1978 Black Hills	PP	M		109	90	50
		H		77		50
1979 Colorado Front Range	PP	Contiguous		175	90	20

TABLE 2. Cost per plot for photo interpretation and ground measurements (in person hours)

		Photo Plot Size			Ground Plot Size
Host Species	Stratum	40 Acres	62.5 Acres	90 Acres	2.5 Acres
LP	L	0.39	0.53	0.73	10.60
	M	0.48	0.66	0.85	12.40
	H	1.04	1.11	1.40	13.10
PP	L	0.33	0.39	0.49	10.60
	M	0.40	0.51	0.63	12.40
	H	0.65	0.81	1.01	13.10

The percent standard error of the estimate is computed as 100 v(y)/y. Formulas for computing variance components are given in the appendix.

A computer program used iterative methods to solve for a fixed percent standard error while varying the Stage 2 allocation for numerous values of n and for 1 = 1, 2, and 3. For each solution, the cost was computed with the formula:

$$C = n\,C_{PI} + ml\,C_G$$

in which

C = total cost for this allocation

C_{PI} = cost of interpreting a photo plot

C_G = cost of measuring a ground plot

the minimum cost was then selected from the possible combinations of n, m, 1.

In order to combine strata, weights were used. The weight for each stratum was the ratio of the expected number of plots that would have fallen in the stratum had no stratification been imposed and the actual number of plots that were sampled in the stratum in the previous survey. The variance component for stage i, v_i, was computed using the formula:

$$v_i = \sum_{j=1} w_{ij} v_{ij}$$

in which

v_i = the variance component for stage i

k = the number of strata

w_{ij} = the weight for stage i, stratum j

v_{ij} = the variance component for stage i, stratum j

Likewise, since the costs were dependent upon whether the plots were from the light, medium, or heavy stratum, weighted costs were used in the combined strata analysis.

RESULTS

The results are summarized (Tables 3 through 6) to provide a readily usable tool with which to better plan future surveys. For each host species, those sampling allocations which yielded the minimum cost for the various levels of precision are provided. In addition, the costs for allocations other than the optimum are presented. This is for planning purposes. Often, cost is not the

TABLE 3. Optimal Sampling Allocations for the 1978 Beaverhead-Gallatin Survey in Lodgepole Pine.

STRATUM	% Std. Error	One Ground Plot Per Stage 2 Plot			Two Ground Plots Per Stage 2 Plot			Three Ground Plots Per Stage 2 Plot		
		No. of PI Plots	No. of PI Sub-Plots	Cost ($)	No. of PI Plots	No. of PI Sub-Plots	Cost ($)	No. of PI Plots	No. of PI Sub-Plots	Cost ($)
L	10%	300	21	339	450	14	471	400	13	569
M		200	19	331	250	12	416	250	10	491
H		200	18	574	250	16	679	250	13	770
TOTAL		700	58	1244	950	42	1566	900	36	1830
NO STRATIFICATION		300	28	517	350	18	626	350	15	733
L	15%	150	9	154	150	7	207	200	6	258
M		80	8	137	60	6	177	60	5	215
H		100	13	274	150	7	339	100	7	379
TOTAL		330	30	565	360	20	723	360	18	852
NO STRATIFICATION		140	12	241	140	8	282	150	7	331
L	20%	80	5	84	70	4	112	70	3	123
M		50	4	73	40	3	93	40	3	131
H		60	7	154	60	4	179	70	4	203
TOTAL		190	16	311	170	11	384	180	10	457
NO STRATIFICATION		80	7	130	90	4	158	90	4	184
L	25%	50	3	51	50	2	62	50	2	83
M		20	3	47	30	2	64	20	2	84
H		40	4	100	40	3	115	40	2	130
TOTAL		110	10	198	120	7	241	110	6	297
NO STRATIFICATION		50	5	84	50	3	102	50	2	120

TABLE 4. Optimal Sampling Allocations for the 1979 Montana Survey in Lodgepole Pine.

STRATUM	% Std. Error	One Ground Plot Per Stage 2 Plot			Two Ground Plots Per Stage 2 Plot			Three Ground Plots Per Stage 2 Plot		
		No. of PI Plots	No. of PI Sub-Plots	Cost ($)	No. of PI Plots	No. of PI Sub-Plots	Cost ($)	No. of PI Plots	No. of PI Sub-Plots	Cost ($)
L	10%	500	110	1360	500	70	1678	500	56	1975
M		600	177	2480	700	104	2913	700	81	3339
H		400	40	944	400	34	1318	500	29	1676
TOTAL		1500	327	4784	1600	208	5909	1700	166	6990
NO STRATIFICATION		900	114	1904	1100	77	2519	1200	66	3113
L	15%	250	48	606	300	30	752	250	25	892
M		250	81	1121	300	47	1315	300	37	1508
H		150	20	421	200	15	594	200	14	754
TOTAL		650	149	2148	800	92	2661	750	76	3154
NO STRATIFICATION		400	51	852	500	34	1127	500	30	1392
L	20%	110	28	339	80	18	413	100	14	484
M		50	45	634	150	27	744	200	20	855
H		60	11	238	100	9	335	120	8	426
TOTAL		220	84	1211	330	54	1492	420	42	1765
NO STRATIFICATION		250	28	480	250	20	636	300	17	785
L	25%	70	18	218	70	11	260	90	9	321
M		100	29	407	120	17	479	150	13	555
H		60	7	152	70	5	215	80	5	275
TOTAL		230	54	777	260	33	954	320	27	1151
NO STRATIFICATION		150	18	307	200	12	408	200	11	503

TABLE 5. Optimal Sampling Allocations for the 1978 Black Hills Survey in Ponderosa Pine*

STRATUM	% Std. Error	One Ground Plot Per Stage 2 Plot			Two Ground Plots Per Stage 2 Plot			Three Ground Plots Per Stage 2 Plot		
		No. of PI Plots	No. of PI Sub-Plots	Cost ($)	No. of PI Plots	No. of PI Sub-Plots	Cost ($)	No. of PI Plots	No. of PI Sub-Plots	Cost ($)
M	10%	1400	62	1646	1400	37	1797	1500	27	1944
H		400	52	1084	400	32	1246	500	23	1407
TOTAL		1800	114	2730	1800	69	3043	2000	50	3351
M	15%	700	35	870	800	18	953	800	14	1027
H		200	25	530	200	15	607	200	12	687
TOTAL		900	60	1400	1000	33	1560	1050	26	1714
M	20%	400	22	528	500	11	577	500	8	621
H		110	15	309	120	9	355	130	7	397
TOTAL		510	37	837	620	20	932	630	15	1018
M	25%	300	13	347	300	8	378	300	6	411
H		70	10	201	80	6	230	80	5	260
TOTAL		370	23	548	380	14	608	380	11	671

*Strata were not combined since no data were available for the light condition.

TABLE 6. Optimal Sampling Allocations for the 1979 Colorado Survey in Ponderosa Pine.

STRATUM	% Std. Error	One Ground Plot Per Stage 2 Plot			Two Ground Plots Per Stage 2 Plot			Three Ground Plots Per Stage 2 Plot		
		No. of PI Plots	No. of PI Sub-Plots	Cost ($)	No. of PI Plots	No. of PI Sub-Plots	Cost ($)	No. of PI Plots	No. of PI Sub-Plots	Cost ($)
	10%	1400	352	5240	1600	228	6654	1800	186	8044
	15%	600	165	2421	800	104	3079	900	85	3725
	20%	400	91	1378	400	61	1763	500	49	2135
	25%	200	62	894	200	35	894	300	38	1130

only consideration in allocation of resources. Scheduling, amount of training required, travel restrictions, etc., often play a part. Therefore, it is important to be able to know how much will be sacrificed in having a design which is suboptimum in some respect.

Some of the results were consistent over all of the previous survey data sets. In no case was stratification beneficial. This is probably because in variable probability sampling the measurement upon which the variance depends is the ration of the next stage measurement to the previous stage measurement. This would not be necessarily more homogeneous within strata defined by intensities of mortality. In every case, sampling with only one ground plot per photoplot in the final stage was best.

Results for ponderosa pine are somewhat limited in that for the Black Hills, data from only two strata were available and for the Colorado Front Range survey, only one stratum was usable. The results are summarized in Tables 5 and 6. Since data were incomplete, we were not able to do the analysis for no stratification.

DISCUSSION

The variances in any two surveys will not be the same. Much depends upon the geographic location of the survey, the quality of the photos, the skill of the interpreters, and the inherent variability in the population being surveyed. Consequently, the results presented here should be used conservatively as a guideline, not as an absolute rule.

It is felt that the results derived from these analyses are highly dependent on the cost information used. If more precise answers are to be obtained, more effort should be directed in the future to obtain and maintain cost data. For each survey, a good estimate of costs could be obtained if the total person hours spent doing photo interpretation and the total number of photos interpreted were tallied, as well as the total person hours spent doing ground work plus the number of plots measured on the gorund were recorded.

It is not necessary to take more than one plot in the final stage of sampling in order to estimate the standard error of the estimate of total for any single survey; however, it is impossible to evaluate the variance component in the final stage for the optimization of future surveys if only one plot is sampled. For this reason, it is often desirable to consider only those designs with at least two ground plots per photo plot in the final stage even though this may not be the most cost effective. Also, in choosing a viable alternative for a particular survey, some provision should be made to allow for missing or unusable data. This is another reason why it may be better to use more than one ground plot per cell in the final sampling stage.

REFERENCES

Bennett, D.D., and W.E. Bousfield. 1979. A pilot survey to measure annual mortality caused by the mountain pine beetle in lodgepole pine on the Beaverhead and Gallatin National Forests. USDA For. Serv., Northern Region, Rep. No. 79-20. 13 pp.

Bennett, D.D., and W.E. Bousfield, M.D. McGregor, and K.E. Gibson. 1980. Evaluation of multistage sampling techniques to measure lodgepole pine mortality caused by mountain pine beetle in Montana, 1979. USDA For. Serv., Northern Region, Rep. No. 80-13. 11 pp.

Ciesla, W.M., and F.M. Yasinski. 1980. Forest insect and disease information systems (FIDIS) implementation plan. USDA For. Serv., FIDM/MAG Rep. No. 80-3. 10 pp.

Hosteller, B.B., and R.W. Young. 1979. Estimation procedures for determining annual tree mortality caused by the mountain pine beetle. USDA For. Serv., Rocky Mtn. Region, Tech. Rep. R2-20. 25 pp.

Klein, W.H., D.D. Bennett, R.W. Young. 1979. A pilot survey to measure annual mortality of lodgepole pine caused by the mountain pine beetle. USDA For.Serv., FIDM/MAG Rep. No. 78-4. 15 pp.

Lister, C.K., and R.W. Young. 1981. 1979 Colorado forest range mountain pine beetle survey. USDA For.Serv., Rocky Mtn. Region, Tech. Rep. R2-22. 19 pp.

APPENDIX

The variance components of the unbiased estimator for the number of dead trees are defined in this section. Let $V(Y_1)$ be the variance component associated with the first stage sampling, then $V(Y_1)$ is as follows:

$$V(Y_1) = \frac{N(N-n)}{n} \quad \frac{\sum_{i=1}^{N} x_i^2 - N'\bar{x}^2}{n-1}$$

Let $V(Y_2)$ be the variance component associated with the second stage sampling, the $V(Y_2)$ is as follows:

$$V(Y_2) = \frac{1}{m}\left(\frac{N}{N'}\right)^2 \quad \frac{\sum_{J=1}^{M'} \frac{v_j^2}{P_j^2} - M'\hat{V}^2}{M'-1}$$

Let $V(Y_3)$ be the variance component associated with the third stage sampling, the $V(Y_3)$ is as follows:

$$V(Y_3) = \left(\frac{N}{N'}\right)^2 \frac{1}{m\ell}\left\{\frac{N'}{M'}\sum_{J=1}^{M'} \frac{1}{P_j} \quad \frac{\sum_{k=1}^{L'} \frac{v_{jk}^2}{P_{jk}^2} - L'\hat{V}_j^2}{L'-1}\right\}$$

The definitions for the variables applicable to the above formulas as follows:

N = total possible number of PI plots

N' = total number of PI plots taken

x_i = total PI for plot i.

x = mean of Pi values for all photos

v_{jk} = ground measurement for photo j and ground plot k

p_{jk} = probability of ground plot k within photo plot j

V_j = $V_j = \bar{V}_j = \frac{1}{L'}\sum_{k=1}^{L'} \frac{v_{jk}}{P_{jk}}$

L' = number of ground plots samled per photo plot

L = total possible number of ground plots/photo plot

p_j = probability for selecting photo plot j

M' = number of photo plots sub-sampled during second stage in original survey

m = number of photo plots to sub-sample during second stage for next time (parameter to be optimized)

l = number of ground plots/photo plot to sample next time (parameter to be optimized)

n = number of photos in first stage next time (parameter to be optimized)